

Лабораторная работа № 2

«Критерии согласия и однородности выборок»

студента Яковлева Андрея группы Б20-504. Дата сдачи: 11.11.2022 _

Ведущий преподаватель: Трофимов А.Г. оценка: _____ подпись: _____

Вариант № 12

Цель работы: изучение функций Statistics and Machine Learning Toolbox™ MATLAB / Python SciPy.stats для проверки критериев согласия (*goodness-of-fit tests*) и однородности выборок.

1. Исходные данные

Характеристики наблюдаемой случайной величины X :

Распределение	Параметры	Математическое ожидание, m	Дисперсия, σ^2
$N(\mu, \sigma^2)$	$\mu = 5$ $\sigma^2 = 3$	5	3

Объём выборки $n_1 = 250$

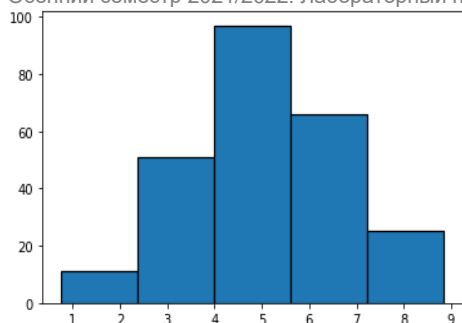
Примечание: для генерации случайных чисел использовать функции **rand**, **randn**, **chi2rnd** (scipy.stats: **uniform.rvs**, **norm.rvs**, **chi2.rvs**)

Выборочные характеристики:

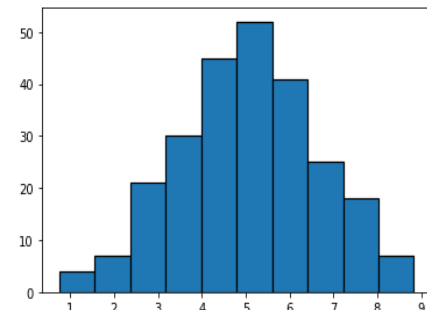
Среднее, \bar{x}	Оценка дисперсии, s^2	Оценка с.к.о., s
5.060	2.433	1.560

2. Визуальное представление выборки

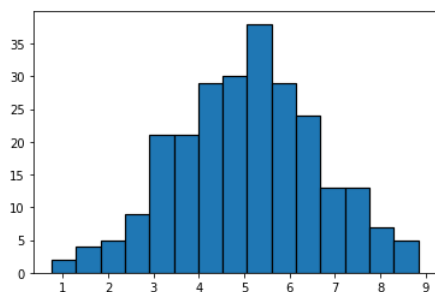
Гистограммы частот:



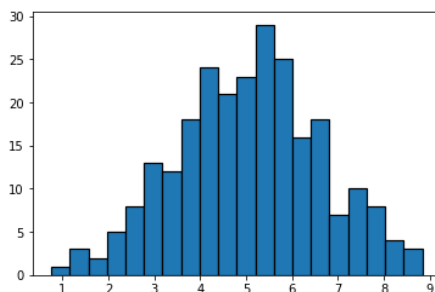
$NBins = 5$



$NBins = 10$



$NBins = 15$



$NBins = 20$

Примечание: для построения гистограмм использовать функцию **hist** (**scipy.stats**: **histogram**; **matplotlib.pyplot**: **hist**)

3. Критерий хи-квадрат

a) Статистическая гипотеза: $H_0: X \sim N(m, \sigma)$

Число интервалов группировки	Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.1$	Ошибка стат. решения
5	1.117	0.856	H_0 принята	Нет
10	3.144	0.257	H_0 принята	Нет
15	6.354	0.206	H_0 принята	Нет
30	24.800	0.829	H_0 принята	нет

б) Статистическая гипотеза: $H_0: X \sim R$

Число интервалов группировки	Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.1$	Ошибка стат. решения
5	21.937	$3.4 \cdot 10^{-5}$	H_0 отвергнута	Нет
10	15.736	0.006	H_0 отвергнута	Нет
15	18.284	0.02	H_0 отвергнута	Нет
30	39.874	0.003	H_0 отвергнута	нет

в) Статистическая гипотеза: $H_0: X \sim \chi^2(5)$

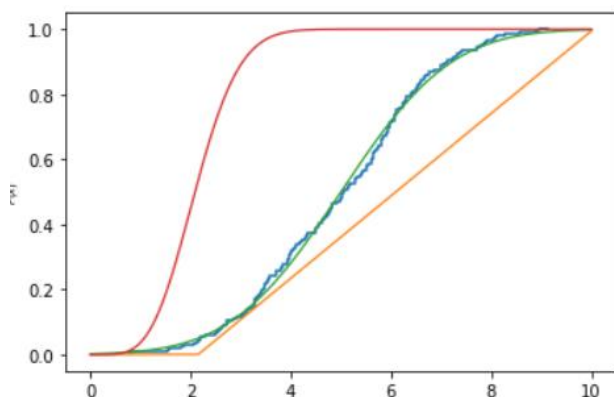
Число интервалов группировки	Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.1$	Ошибка стат. решения
5	96.150	0.00	H_0 отвергнута	Нет
10	106.738	0.00	H_0 отвергнута	Нет
15	115.760	0.00	H_0 отвергнута	Нет
30	137.022	0.00	H_0 отвергнута	нет

Примечание: при расчетах использовать функции **chi2gof**, **fitdist** (**scipy.stats**: **histogram**, **chisquare**)

4. Критерий Колмогорова

Статистическая гипотеза, H_0	Выборочное значение статистики критерия	p -value	Статистическое решение при $\alpha = 0.1$	Ошибка стат. решения
$X \sim N(m, \sigma)$	0.029	0.983	H_0 принята	нет
$X \sim R$	0.069	0.179	H_0 принята	Ошибка второго рода
$X \sim \chi^2(5)$	0.222	0.000	H_0 отвергнута	нет

Примечание: при расчетах использовать функции **kstest**, **lillietest**, **fitdist** (**scipy.stats**: **kstest**)



Эмпирическая и теоретические функции распределения

Примечание: для построения графиков использовать функции **ecdf**, **cdf** (**scipy.stats**: **uniform.cdf**, **norm.cdf**, **chi2.cdf**; **statsmodels.distributions.empirical_distribution**: **ECDF**)

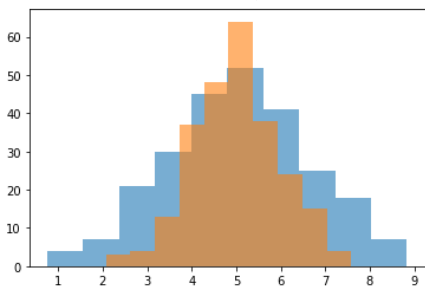
5. Двухвыборочные критерии

Характеристики наблюдаемой случайной величины Y :

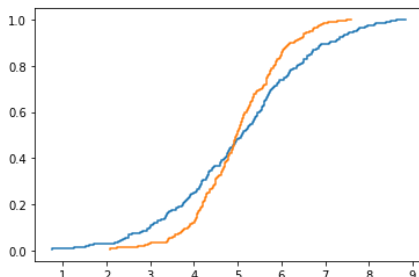
Распределение	Параметры	Математическое ожидание	Дисперсия
$N(\mu, \sigma^2)$	$\mu = 5$ $\sigma^2 = 1$	5	1

Объём выборки $n_2 = 250$

HIST: X, Y



ECDF: X, Y



Критерий	Стат. гипотеза, H_0	Выборочное значение статистики критерия	p -value	Стат. решение при $\alpha = 0.1$	Ошибка стат. решения
Chi-squared	$H_0: F_X(\xi) = F_Y(\xi)$	14.397	0.109	H_0 принята	Нет
KS-test	$H_0: F_X(\xi) = F_Y(\xi)$	0.136	0.020	H_0 отвергнута	Ошибка первого рода
Sign test	$H_0: F_X(\xi) = F_Y(\xi)$	2.000	0.850	H_0 принята	Нет
U-test	$H_0: F_X(\xi) = F_Y(\xi)$	0.417	0.677	H_0 принята	Нет

Примечание: при расчетах использовать функции **chi2gof**, **kstest2**, **signtest**, **ranksum** (scipy.stats: **chisquare**, **ks_2samp**; statsmodels.stats.descriptivestats: **sign_test**, **ranksums**)